



(S//SI) SID and RAD to Develop Lexicons in Five Languages

FROM: Dr. David Savignac
CAMT Technical Director (S21212)
Run Date: 04/09/2004

FROM: Dr. David Savignac
CAMT Technical Director (S21212)

(S//SI) Thanks to \$500K of Intelligence Community funding forwarded by NSA's Counter-Biological Weapons Program, SID's Center for Applied Machine Translation (CAMT) has begun to create lexicons covering biological weapons terminology in five critical languages: Russian, Chinese, Korean, Arabic, and Farsi. NSA's Counter-BW Program is jointly run by the Research Associate Directorate (RAD) and the Office of Proliferation and Arms Control (S2G), and is funded by the Intelligence Community's Intelligence Technology Innovation Center (ITIC).

(S//SI) ITIC had originally provided the funds to NSA to improve machine translation algorithms for the biological weapons target, but in the course of discussions between [REDACTED] (RAD) and Dr. David Savignac (CAMT), it became clear that ITIC's requirement could be best met by having the CAMT create appropriate lexicons in the target languages. ITIC concurred to this modified approach and was pleased to learn that the CAMT believed that it could stretch the money to cover lexicon development in more than the original five languages.

(S//SI) Once completed, the new lexical information will be immediately integrated into the CAMT's CyberTrans machine translation software to improve automated translation capabilities for this extremely important domain. The lexicons will also be provided to commercial machine translation vendors (Systran, Language Weaver and possibly others) for inclusion in software which they supply to the US Government; they will also be released to field sites, training institutions, and the relevant target offices.

(C) The CAMT's CyberTrans software is used throughout the Intelligence Community, the Department of Defense, and 2nd and 3rd Parties to translate foreign language text from some sixty languages into English. The biological warfare lexicon project will proceed under the leadership of Dr. Carl Rubino of the CAMT, who is presently overseeing the creation of WMD lexicons in Arabic, Turkish and Uzbek.

POC: Dr. Carl Rubino, CAMT/S21212, [REDACTED]@nsa, [REDACTED]

"(U//FOUO) SIDtoday articles may not be republished or reposted outside NSANet without the consent of S0121 ([DL sid comms](#))."